PB90-242306

Rationale for the Assessment of Errors in the Sampling of Soils

(U.S.) Environmental Monitoring Systems Lab., Las Vegas, NV

Jul 90

U.S. Department of Commerce
National Technical Information Service
NTIS

PB90-242306

EPA/600/4-90/013
July 1990

A RATIONALE FOR THE ASSESSMENT OF ERRORS IN THE SAMPLING OF SOILS

by

J. Jeffrey van Ee and Louis J. Blume
Exposure Assessment Division
Environmental Monitoring Systems Laboratory
Las Vegas, Nevada 89193

and

Thomas H. Starks
Environmental Research Center
University of Nevada - Las Vegas
Las Vegas, Nevada 89154

Environmental Monitoring Systems Laboratory
Office of Research and Development
Las Vegas, Nevada 89193-3478

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL
INFORMATION SERVICE
SPRINGFIELD, VA 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before complet

1. REPORT NO. EPA/600/4-90/013	2.
3. PB90-242306	
4. TITLE AND SUBTITLE A RATIONALE FOR THE ASSESSMENT OF ERRORS IN THE SAMPLING OF SOILS	5. REPORT DATE July 1990
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) J. Jeffrey van Ee, Louis J. Blume Thomas H. Starks	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Environmental Protection Agency, P.O. Box 93478 Las Vegas, NV 89193-3478; Environmental Research Center, UNLV, Las Vegas, NV 89154	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring Systems Laboratory - LV, NV Office of Research and Development U.S. Environmental Protection Agency Las Vegas, NV 89193-3478	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE EPA/600/07
15. SUPPLEMENTARY NOTES	

16. ABSTRACT The sampling of soils in RCRA and Superfund monitoring programs requires associated quality assurance programs. One objective of any quality assurance program is to assess and document the quality of the study data to ensure that it satisfies the needs of the users. The purpose of this document is to describe the mature and function of certain quality assurance samples in the assessment and documentation of bias and precision in sampling studies of inorganic pollutant concentrations in soils. A foundation is provided for answering two basic questions:

How many, and what type of, quality assurance (or, to be more specific, quality assessment) samples are required to assess the quality of data in a fields sampling effort?

How can the information from the quality assessment samples be used to identify and control the principal sources of error and uncertainty in the measurement process?

This document has been developed to provide people who plan, implement, or oversee RCRA or Superfund soil sampling studies with information on quality assessment samples so that they will have a better basis for decisions concerning the employment of such samples in their quality assurance programs.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)* UNCLASSIFIED	21. NO. OF PAGES 65
	20. SECURITY CLASS *(This page)* UNCLASSIFIED	22. PRICE A04

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

The information in this document has been funded wholly or in part by the United States Environmental Protection Agency under Cooperative Agreement CR814701-01 . It has been subject to the Agency's peer and administrative review, and it has been approved for publication as an EPA document.

RECOMMENDATIONS FOR FUTURE WORK 45

REFERENCES 46

APPENDIX A
KEY COMPONENTS 48
Training 48
Pilot Study 48
Audits 48
Documentation 49

APPENDIX B 50
DEFINITIONS 51
Quality Assurance 51
Quality Control 51
Quality Assessment 51
Soil 51
Standard Additions 51

APPENDIX C 52
Soil Sampling Methods Table 53

APPENDIX D 53
The Lognormal Distribution and Logarithmic Transformations 54

APPENDIX E 54
Non-blind Quality Assessment Samples in the Contract Laboratory Program 55

APPENDIX F 55
Upper Confidence Limits for the Variance, σ^2, as a Function of the Number of Degrees of Freedom for the Variance Estimate, s^2. 57

LIST OF FIGURES

Figure		Page
1.	Quality Assessment Samples	24
2.	Alternative Quality Assessment Sample Design	37
3.	Scatter Plot of QA Data from ASSESS	43
4.	Error Plot from ASSESS	44

TABLE OF CONTENTS

	Page
LIST OF FIGURES	v
LIST OF TABLES	vi
ACKNOWLEDGMENT	vii
INTRODUCTION	1
Purpose	1
BACKGROUND	3
Sources of Error	3
Representative Sampling	5
Data Quality Levels for Analytical Measurements in Superfund	7
NUMBER OF QUALITY ASSESSMENT SAMPLES	9
Background	9
Confidence Levels for the Assessment of Measurement Variability	10
Quantitative Assessments of Bias	11
Recommendations for the Assessment and Control of Bias and Measurement Variability	12
DEFINITIONS AND TERMS	15
Quality Assessment Samples	15
Split Samples	19
Spiked Samples	19
Blanks	19
Batch	19
AVAILABILITY OF QUALITY ASSESSMENT SAMPLES	21
A RATIONALE FOR ASSESSING ERRORS	22
Quality Assessment Sample Design	22
Field Duplicates	25
Preparation Splits	26
Field Evaluation Samples	27
External Laboratory Evaluation Samples	28
Preparation Rinsate Blanks	29
Field Rinsate Blanks	30
Equations for Estimating Bias and Precision	31
AN ALTERNATIVE QA DESIGN THAT DOES NOT EMPLOY FES AND ELES	34
EXAMPLE OF QUALITY ASSESSMENT	38
QA/QC from a Pilot Study	38
Post Pilot Study Data Analysis	39
Development of the QA Plan for the Primary Study	44

LIST OF TABLES

Table		Page
1.	Sources of Bias in Soil Sampling Studies	4
2.	Sources of Variability in Soil Sampling Studies	5
3.	Some 95 Percent Confidence Intervals for Variance	10
4.	Types of Quality Assessment Samples or Procedures	17-18
5.	Equations for Determining Precision and Bias	32-33
6.	Equations for Determining Precision without FES and ELES	35
7.	Quality Assessment Data	41
8.	Transformed Quality Assessment Data	42

ACKNOWLEDGMENT

During the initial development of this document, Kenneth W. Brown and Daniel T. Heggem of the Exposure Assessment Division contributed their knowledge, words, and constructive suggestions to the development of the first drafts. Later, Ann M. Pitchford of the Exposure Assessment Division made further constructive changes to the developing document. The document was then reviewed by a variety of people with a variety of backgrounds since the guidance provided in this document is meant to apply to an entire soils study which involves a diverse group of people. The resulting document is meant to serve as an initial "roadmap" for people to use in assessing the errors and uncertainties in a soils study. With time the importance of the guidance provided by this document is expected to become greater as more experience is gained in applying the rationale to greater numbers of hazardous waste site investigations. Further revisions are anticipated, but those revisions should not detract from the significant contributions that were made in the development of this document to the assessment of <u>total</u> quality of data from a soils investigation.

INTRODUCTION

Considerable guidance has been provided on the importance of quality assurance (QA), quality control (QC), and quality assessment procedures for assessing and minimizing errors in environmental studies. QA/QC terms, such as quality assurance project plans and program plans, are becoming a part of the vocabulary for remedial project managers (RPM's). Establishment of data quality objectives (DQOs) early in the process of a site investigation has been stressed in EPA QA/QC guidance documents. Quality assessment practices, such as the use of duplicate, split, spiked, and reference samples, are becoming widely accepted as important means for assessing errors in measurement processes. Despite the existence of numerous and various forms of guidance for hazardous waste site investigations, there have been no clear, concise, well-defined strategies for precisely how recommended QA/QC practices can be utilized.

Purpose

The purpose of this document is to provide a foundation for answering two basic questions:

> How many, and what type, of samples are required to assess the quality of data in a field sampling effort? (quality assessment samples)
>
> How can the information from the quality assessment samples be used to identify and control sources of error and uncertainties in the measurement process?

This document expands upon the guidance for quality control samples for field sampling as contained in Appendix C of EPA's Data Quality Objectives for Remedial Response Activities - Development Process (9). This report outlines, in greater detail, strategies for how errors may be assessed and minimized in the sampling of soils with emphasis on inorganic contaminants.

Basic guidance for soil sampling, which includes a discussion of basic principles, may be found in EPA's Revised Soil Sampling Quality Assurance Users Guide (15). The Users Guide is intended to be revised on a periodic basis. It is anticipated that some of the guidance provided in this document will eventually be incorporated into the Users Guide.

The primary audience for this document is assumed to be RPM's who have a concern about the quality of the data being collected at Superfund sites but have little time to understand the complexities of the processes used to assess the quality of data from the total measurement process. The approach outlined in this document for assessing errors in the field sampling of

inorganics in soils may be transferrable, with modification, to other contaminants in other media.

The example offered at the end of this document illustrates the planning process for determining a reasonable number of quality assessment samples. The example also demonstrates how the information from the process may be used to document the quality of the measurement data, and how this data may be used to make adjustments to the monitoring program.

BACKGROUND

Superfund and RCRA site investigations are complicated by: the variety of media being investigated, an assortment of methods, the diversity of people, the variety of contaminants, and the numerous risks and effects to human health and the environment. Many phases exist in Superfund site investigations. An initial phase, generally described as a "preliminary investigation," consists of collecting and reviewing existing data and data from limited measurements using practically any available method. The next phase, generally described as "site characterization," uses selected methods and prescribed procedures to characterize the magnitude and areal extent of the contamination. Final phases include an examination of remedial actions, which involve an examination of treatment technologies, and monitoring to assess the degree of cleanup at a site. A final phase may require long-term monitoring to substantiate that no new or additional threats occur to human health and the environment. Throughout Superfund site investigations QA/QC procedures change as data quality objectives vary and different phases occur.

Sources of Error

In many of the phases of Superfund and RCRA site investigations, errors and uncertainties occur. During the measurement process, random errors will be induced from: sampling; handling, transportation and preparation of the samples for shipment to the laboratory; taking a subsample from the field sample and preparing the subsample for analysis at the laboratory, and analysis of the sample at the laboratory (including data handling errors). The magnitude of these errors can be expected to vary during the measurement process and make it more difficult to determine the natural variability of contaminants in the environment. Errors introduced in the interpretation and analysis of data are not considered in this document.

Typically, errors in the taking of field samples are much greater than preparation, handling, analytical, and data analysis errors; yet, most of the resources in sampling studies have been devoted to assessing and mitigating laboratory errors. It may be that those errors have traditionally been the easiest to identify, assess and control. This document adopts the approaches used in the laboratory, e.g. the use of duplicate, split, spiked, evaluation and calibration samples, to identify, assess and control the errors in the sampling of soils.

Systematic errors, termed bias (B), can accumulate during a measurement process. Bias may result from: faults in sampling design, sampling procedure, analytical procedure, contamination, losses, interactions with containers, deteriorations,

displacement of phase or chemical equilibria, and inaccurate instrument calibrations. (Table 1) Bias causes the mean value of the sample data to be either consistently higher or consistently lower than the "true" mean value. Laboratories usually introduce various quality control samples into their sample load to detect possible bias. Bias in soil sampling is difficult to detect. Components of bias can be discovered by the technique described as standard additions or by using evaluation samples. On the other hand, it is difficult to demonstrate that bias is not present because an apparent lack of bias may be the result of an inability to measure it rather than its actual absence.

Table 1. Sources of Bias in Soil Sampling Studies

B_s = Measurement bias introduced in sample collection not caused by contamination
B_{sc} = Measurement bias introduced in sample collection caused by contamination
B_h = Measurement bias introduced in handling and preparation not caused by contamination
B_{hc} = Measurement bias introduced in handling and preparation caused by contamination
B_{ss} = Measurement bias introduced in subsampling not caused by contamination
B_{ssc} = Measurement bias introduced in subsampling caused by contamination
B_a = Measurement bias introduced in the laboratory analytical process not caused by contamination
B_{ac} = Measurement bias introduced in the laboratory analytical process caused by contamination
B_n = Total measurement bias

NOTE: It is necessary to realize that biases, other than contamination biases in the measurement of a sample, will often be dependent on the original concentration of the contaminant being measured and on the sample matrix. Biases caused by contamination are listed separately because some QA samples, such as rinsate samples, detect only contamination bias.

Also, variability occurs in the measurement process from the heterogeneity of the soil and random errors throughout the measurement process. The variability caused by any type of random error is frequently described quantitatively by the variance, σ^2, of the random error, or by the positive square root, the standard deviation, σ, of the random error. Variances of independent random errors are additive in that the variance of the sum of errors is the sum of the variances of the individual errors (Table 2). Other quantifications of variability do not have this useful, additive property.

Table 2. Sources of Variability in Soil Sampling Studies

$$\sigma_t^2 = \sigma_m^2 + \sigma_p^2$$

where σ_t = total variability
σ_m = measurement variability
σ_p = population variability

$$\sigma_m^2 = \sigma_s^2 + \sigma_h^2 + \sigma_{ss}^2 + \sigma_a^2 + \sigma_b^2$$

where σ_s = sampling variability (standard deviation)
σ_h = handling, transportation and preparation variability
σ_{ss} = preparation variability (subsampling variability)
σ_a = laboratory analytical variability
σ_b = between batch variability

NOTE: It is assumed that the data are normally distributed or that a normalizing data transformation has been performed.

Biases and variability can accumulate during a measurement process to the point where the data are unsuitable for meeting the objectives of the study. Often at the end of, but preferably during the planning of a study, a question arises as to whether the data are acceptable in terms of accuracy, precision, representativeness, and completeness, i.e. DQOs. Quality assessments, i.e., systematic investigations of the measurement process, can be performed to try to assess and identify the extent of biases and variability in the measurement process and to determine whether the DQOs are being met.

Representative Sampling

Soils are extremely complex and variable which necessitates a multitude of sampling methods. The sample collector must select methods that best accommodate specific sampling needs, and that satisfy the stated sampling objectives. In addition, the sample collector is responsible for providing the appropriate sample for laboratory analysis. A soil sample must satisfy the following:

1. Provide an adequate amount of soil to meet analytical requirements and be of sufficiently large volume as to keep short range variability reasonably small,
2. Provide material < 2 mm in size,
3. Be a member of the population to be evaluated and, when

taken in association with the other samples, be representative of that population.

Deposition of airborne contaminants, especially those recently deposited, is often evident in the surface layer of the soil. Contaminants that have been deposited by liquid spills or by long-term disposition of water soluble materials may be found at depths up to several meters. Plumes emanating from hazardous waste dumps or leaking storage tanks may be found at considerable depths. The methods of sampling each of these may be different; but all make use of one of the following three basic sampling tools: (1) scoops, (2) coring, or (3) augering devices.

Two major considerations must be addressed when selecting a specific sampling tool. These two considerations include soil conditions and the contaminant(s) that are to be analyzed from the collected material. Soil condition can be extremely variable from location to location. For example, soils can be wet or dry, stony, cohesive (e.g., clay) or cohesionless (e.g., sand). Similarly, contaminants are extremely diverse, varying between metals which in most cases are relatively immobile, to highly mobile water soluble substances, to contaminants that are volatile.

Improper use and selection of sampling tools may result in data that are not representative of the soil environment being sampled (See Appendix C). Measurement errors can result from a tool being either inappropriate for the particular task, or improperly used. Results based on previous experience, or from an equivalency test, may be used to evaluate and select the proper tool for a specific sampling objective. Operational measurement errors are identified and assessed by implementing and utilizing a number of field QA samples. The optimal number and timing of QA samples depend in part on the proper soil sampling method being utilized.

A variety of sampling methods may be used to obtain a measurement of inorganics in soil. EPA's Soil Sampling Quality Assurance Users Guide notes that the concentrations measured in an heterogenous medium such as soil are related to the volume of soil sampled and the orientation of the sample within the volume of earth that is being studied. (The term "support" is used to describe this concept.) A RPM not only wants to know the concentration of contaminants, but their location. Frequently, an average concentration of contaminants in the soil is sought and compared against some standard. Depending on the sampling method used, the location of the samples collected, the number of samples taken, and the time the samples were collected, the reported concentrations can vary considerably even when relatively stable contaminants such as inorganics in soil are measured.

The processes involved in collecting representative samples of inorganics in soil can be complicated. (The Soil Sampling Quality Assurance Users Guide should be consulted for further information on these processes.) The problem of measuring the natural variability of contaminants, such as inorganics, in soil and adequately representing the site to be studied is also a problem for traditional QA/QC programs which have emphasized the assessment and minimization of errors and variabilities in the analytical process.

A major problem in obtaining a "representative" sample is the spatial scale chosen for the study. Geostatistical techniques, such as kriging, may also be used to estimate the natural variability of contaminants in soil. A measure of the spread of the distribution of contaminant concentrations about the mean concentration is the population variance, σ_p^2.

Data Quality Levels for Analytical Measurements in Superfund

As many as five different levels of quality have been assigned by EPA in the Superfund program to analytical data. These levels have been generally associated with different phases of a site investigation (9); however, it may prove to be necessary to have all five levels of data quality in any one phase of a site investigation. Levels III and IV involve off-site analytical laboratory measurements with Level IV in the Contract Laboratory Program (CLP) having the most rigorous QA/QC protocols and documentation. Levels III and IV are assumed in this document in the development of a strategy for assessing field sampling errors.

Since the assumption is made in this document that CLP laboratories are involved in the analysis of samples from a soil sampling study, errors and biases from those laboratory measurements are presumed to be small and known. These assumptions allow greater emphasis to be given to the identification of errors and biases in the field sampling rather than in the laboratory analysis.

In a pilot study, within a particular phase of a Superfund site investigation, it may be necessary to utilize Level III and IV analytical levels to identify, assess and reduce errors in field sampling even though these analytical levels might not be needed for every sample and measurement. For example, a field portable x-ray fluorescence instrument, which measures inorganics in soil, is frequently used to identify sampling locations for samples to be sent to the CLP. The data quality from the portable x-ray fluorescence instrument is not classified as being Level III or IV; however, data from the instrument is used to screen samples for subsequent analysis by Level III and IV methods. It may be advantagous to compare the performance of the field-screening, portable x-ray fluorescence instrument against

more rigorous, well characterized laboratory analytical methods even though the level of data quality desired from each method is different.

The assessment of errors from non-conventional "sampling" and analytical methods, such as the portable x-ray fluorescence instrument, are not specifically addressed in this document.

Two important factors must be considered in the collection of environmental data. These are the probability that the collected data will yield the correct assessment or solution for an environmental problem and the costs. The strategy developed in this document recognizes that these important factors must be considered in the implementation of QA/QC measures in the sampling of soils for inorganics.

NUMBER OF QUALITY ASSESSMENT SAMPLES

Background

A key question for a RPM is: how many samples must be collected to adequately characterize the site? A question for a QA/QC officer is: how many quality assessment samples must be taken to adequately characterize the errors and uncertainties in the measurement process? The timing and type of those quality assessment samples in the measurement process determines the type of information that is obtained. The number of quality assessment samples is determined by the available resources and the degree to which investigators need to be sure that they have adequately characterized the measurement process. The simplest case is when one method, one sampling crew, and one laboratory are used to analyze the soil samples. A more difficult, and probably more typical, case is when more than one batch of samples are either collected or analyzed at various times or by various laboratories.

The percentage of the total monitoring effort allocated to QA/QC activities will depend on many factors including the size of the project, the available knowledge concerning sampling and analytical procedures, the relationship of risk to human health and the environment at various pollutant concentrations, the nearness of action levels to method detection limits, and the natural variability and distribution of the contaminants. Typically the smaller the project, the larger will be the proportion of cost allocated to QA/QC. New, untried procedures will typically require pilot-study runs and additional training for personnel. If the action level is near the method detection limit, there will be little room for error in the measurements, and the QA/QC effort may have to be large to assure that measurement errors are kept small. If the natural variability of the contaminants is relatively large, it may be necessary to collect more samples rather than collect more quality assessment samples. One should not specify a certain percentage of a project's costs be allocated to QA/QC without considering the above factors.

Previous EPA guidance for the number of quality assessment samples has been one for every 20 field samples (9). However, such rules of thumb are oversimplifications and should be treated with great caution. A better approach is to determine how each type of QA sample is to be employed and then determine the number for that type based on the use. For example, field duplicates, i.e., duplicate samples at the same location, are used to estimate the combined variance contribution of several sources of variation. Hence, the number of field duplicates to be obtained in the study should be dictated by how precise one wants that estimate of the total measurement variance to be.

Confidence Levels for the Assessment of Measurement Variability

The precision of an estimate, s^2, of the "true" variance, σ^2, depends on the number of degrees of freedom for the estimate which is directly related to the number of quality assessment samples. Table 3 gives the 95% confidence intervals for various numbers of degrees of freedom, based on an assumption that the data are, or have been transformed to, normally distributed data. Methods for obtaining such confidence intervals for any number of degrees of freedom are given in most statistics texts.

Table 3. Some 95 Percent Confidence Intervals for Variance

Degrees of Freedom	Confidence Interval
2	$0.27s^2 \leq \sigma^2 \leq 39.21s^2$
3	$0.32s^2 \leq \sigma^2 \leq 13.89s^2$
4	$0.36s^2 \leq \sigma^2 \leq 8.26s^2$
5	$0.39s^2 \leq \sigma^2 \leq 6.02s^2$
6	$0.42s^2 \leq \sigma^2 \leq 4.84s^2$
7	$0.44s^2 \leq \sigma^2 \leq 4.14s^2$
8	$0.46s^2 \leq \sigma^2 \leq 3.67s^2$
9	$0.47s^2 \leq \sigma^2 \leq 3.33s^2$
10	$0.49s^2 \leq \sigma^2 \leq 3.08s^2$
11	$0.50s^2 \leq \sigma^2 \leq 2.88s^2$
12	$0.52s^2 \leq \sigma^2 \leq 2.73s^2$
13	$0.53s^2 \leq \sigma^2 \leq 2.59s^2$
14	$0.54s^2 \leq \sigma^2 \leq 2.49s^2$
15	$0.54s^2 \leq \sigma^2 \leq 2.40s^2$
16	$0.56s^2 \leq \sigma^2 \leq 2.32s^2$
17	$0.56s^2 \leq \sigma^2 \leq 2.25s^2$
18	$0.57s^2 \leq \sigma^2 \leq 2.19s^2$
19	$0.58s^2 \leq \sigma^2 \leq 2.13s^2$
20	$0.58s^2 \leq \sigma^2 \leq 2.08s^2$
21	$0.59s^2 \leq \sigma^2 \leq 2.04s^2$
22	$0.60s^2 \leq \sigma^2 \leq 2.00s^2$
23	$0.60s^2 \leq \sigma^2 \leq 1.97s^2$
24	$0.61s^2 \leq \sigma^2 \leq 1.94s^2$
25	$0.62s^2 \leq \sigma^2 \leq 1.91s^2$
30	$0.64s^2 \leq \sigma^2 \leq 1.78s^2$
40	$0.67s^2 \leq \sigma^2 \leq 1.64s^2$
50	$0.70s^2 \leq \sigma^2 \leq 1.61s^2$
100	$0.77s^2 \leq \sigma^2 \leq 1.35s^2$

If it is decided that 20 degrees of freedom gives satisfactory precision for the estimate of the total measurement variance, one might equally space 20 field-duplicate samples among the routinely collected field samples so as to have 20 duplicates by the end of the study. (Note: Each pair, field duplicate sample and associated routine sample, provides one

degree of freedom in the estimation of the between-collocated-samples variance.) Alternatively, one might take duplicate samples at a fairly high frequency at the start of the study until 10 duplicate pairs are obtained and then obtain the remaining ten duplicate pairs at a reduced rate over the remainder of the study. This second procedure would allow an early estimate of the variance based on 10 degrees of freedom to determine whether the QA plan is resulting in error variances in the range expected, and the remaining ten pairs would allow the after-study variance estimate to take the entire study into account.

The confidence intervals in Table 3 are called two-sided confidence intervals because they put both upper and lower limits on the value of σ^2. However, in practice, one is particularly interested in the upper limit; that is, one is interested in how much larger than the estimated variance might the true variance be. If the true measurement error variance is seriously underestimated in a pilot study, it may cause one to proceed to an expensive final study with an inadequate protocol. If the measurement error variance is underestimated in a final study it may cause the RPM to put more reliance on the study results than are warranted and may also cause an inadequate protocol to be copied in a following study. For these reasons, some may be more interested in one-sided confidence intervals that provide only an upper limit on the true variance (i.e., since a variance cannot be negative, a one-sided confidence interval is between zero and an upper limit). Such upper limits for one-sided confidence limits are provided in Appendix F for confidence levels of 90, 95 and 99 percent. (Intervals between zero and the upper limits in Table 3 would be 97.5 percent one-sided confidence intervals; that is, for 2 degrees of freedom, one would be 97.5 percent confident that σ^2 is between zero and $39.21s^2$.)

Quantitative Assessments of Bias

Quantitative assessments of bias are complicated by the different types of bias that may be present in a study and the different times when those biases may occur. There may be consistent additive-constant biasing error (e.g., the data handling algorithm might add a constant to all measurements entered into the database). There may be consistent multiplicative biasing error (e.g., a recovery error in the analytical system). There are also random biasing errors of the additive and multiplicative types (e.g., sample taking, recovery, contamination, and calibration errors). In sample taking, the field crew may occasionally have the bottom portion of a soil core drop out of the sampler prior to bagging the core; if concentration decreases with depth, this would be a random biasing error that would increase the expected concentration above "true" value. The rate of recovery of a chemical from samples may depend on the individual matrix properties of the

samples. The random biasing errors are most difficult to detect and to quantify. If contamination occurs in one of 20 samples on average, the use of 20 contamination blanks would have a 36 ($=100[1-0.05]^{20}$) percent chance of not encountering a contamination incident. Random biasing errors contribute to measurement variance as well as the bias of the measurement system. The number of samples required to detect random bias will depend on the distribution of the biasing errors, and this distribution will generally be unknown. A major problem in data analysis is the separation of random biasing errors from random nonbiasing errors. Estimation of the magnitude of bias and its effect on the estimates and decisions, would require many more quality assessment samples than are required for the detection of bias. Bias estimates that are reported in the literature are often only estimates of the analytical bias obtained either as the difference in recovery rate from 100% obtained by the method of standard additions, or the average difference between reported and reference values of performance evaluation samples.

With these considerations in mind, it would seem that the best one can do is to include some bias-detecting quality assessment samples in each batch of routine samples and hope that they will detect bias if it is present. If bias is detected, an effort should be made to eliminate the source of bias, rather than attempt to correct routine-sample measurements for bias.

Non-blind samples, such as the calibration check standards at the analytical laboratory (Appendix E), are used to assess bias in the laboratory and provide a quality control function. That is, if measurements of these check standards differ by too much from their reference values, the instrument is declared "out of control" and is re-calibrated. Then, samples between the last in-control reading and the out-of-control reading may be reanalyzed. The frequency of use of samples of this quality control type should be based on the consequences associated with out-of-control data and the costs of the analyses of these samples versus the costs of re-analyzing routine field samples in out-of-control situations. This frequency of use is usually related to the probability of obtaining an out-of-control situation in the laboratory with the objective being to minimize expenditures of both time and money while obtaining data of quality suitable for the intended end-use of the data.

Recommendations for the Assessment and Control of Bias and Measurement Variability

A two-phased approach is suggested to answer the questions posed in the introduction:

> How many, and what type, of samples are required to assess the quality of data in a field sampling effort? (quality assessment samples)

How can the information from the quality assessment samples be used to identify and control sources of error and uncertainties in the measurement process?

The first phase involves the acquisition of data to estimate total measurement variability and bias. The second phase involves identification of the sources of the bias and variability.

The required number of samples will vary depending on the data quality objectives and available resources. The more quality assessment samples that are used, the better the assessment of measurement errors. Five field-duplicate samples, as demonstrated in Table 3, will yield an estimate of the measurement variability that may be high by a factor of 6 or low by a factor of 2.5 with a confidence level of 95%. As noted earlier, accurate assessments of measurement bias are more complicated. A recommendation of one quality assessment sample for bias per batch would allow for the plotting of bias on a control chart to determine if the measured bias is within acceptable limits. (Bias may be random, constant, or varying.) In either the measurement of bias, or measurement variability, the accumulation of historical data is extremely important in judging the appropriate number of quality assessment samples and the relative importance of that data to the overall objectives of the study.

It must be emphasized that estimates of measurement-error variance components are of little value if they are based on so few degrees of freedom that they may differ from the true variances by large factors (Table 3). Hence, even in pilot studies with few routine samples, it is important to obtain measurement-error component variance estimates that are based on a sufficient number of quality assessment samples (i.e., based on a sufficient number of degrees of freedom) that the user can have some confidence that the large estimates actually represent large variances and small estimates represent small variances. Otherwise, corrective actions taken to improve precision may be a waste of money, and failure to take corrective action may result in the failure of the subsequent study.

Experience indicates that an effective quality assurance program is negated if certain key elements of a sampling effort are not adequately addressed. Those elements are: training, pilot studies, audits and documentation. More detailed discussion of those key components is provided in Appendix A; however, the importance of pilot studies to the overall monitoring effort cannot be stressed enough. The experience and data gained during the pilot study can be extremely important to the success of the larger monitoring effort.

Frequently, time and financial constraints lead to a minimum number of samples being collected in the initial measurement phase of a hazardous waste study. If sufficient historical quality assessment data has not been collected for the selected sampling crews, sampling methods, and analytical laboratories involved in the initial measurement phase, an accurate assessment of performance during the "pilot study" will require a relatively large number of quality assessment samples. However, even with an experienced sampling crew, tested sampling methods, and well-characterized analytical laboratories, unique characteristics of a particular site may require an increased number of quality assessment samples to measure performance against stated data quality objectives.

DEFINITIONS AND TERMS

Problems can often occur when important terms in environmental sampling are not defined, poorly defined, or not well understood. Commonly accepted terms such as soil can have many different definitions depending upon the interests and background of an investigator. Major errors can occur if everyone involved in an investigation has a different understanding of a term.

While the definitions contained in this document may not be universal, it is important in an understanding of the overall process of assessing errors and identifying their sources that the terms be defined early. Important steps and samples that are required for the assessment of variability and bias are defined in the main document. Basic definitions that are less critical to an understanding of the quality assurance process are in Appendix B.

Quality Assessment Samples

Many terms have been used to describe quality assessment samples. (Quality evaluation is also used.) Quality assessment samples are defined as those samples that allow statements to be made concerning the quality of the measurement system. These samples have two primary reasons for utilization. They allow assessment of the quality of the data, and most importantly, they allow for control of data quality to assure that it meets the original objectives.

The objective of this section is to identify the various sample types, define what they are used for, and how they have been previously used in characterizing the measurement process.

Quality assessment samples include samples from three groups, based upon whether they are double-blind, single-blind, or non-blind to the analytical laboratory. Double-blind samples are samples that cannot be distinguished from routine samples by the analytical laboratory. Single-blind samples are samples that can be distinguished from routine samples but are of unknown concentration. Non-blind samples are samples that have a concentration and origin that are known to the analytical laboratory. Some references state that quality control samples are only those that are blind to the analytical laboratory (9). Other documents refer to quality control samples as non-blind to the analytical laboratory (3,13). The intent of categorizing quality assessment samples into these categories, i.e., double-blind, single-blind, and non-blind, is to avoid confusion due to terminology. The key point is that all of the samples discussed here refer to those samples that make some statement about the quality of the measurement system. This discussion will not include samples such as background samples or critical samples

(9) because they are required for characterization of the contamination at a site and not for characterization of the errors in the measurement system.

Table 4 and Appendix E list typical quality assessment samples and describe how measurements of these samples are used in the control of the measurement process and in the evaluation of the quality assurance procedures. To obtain an unbiased measure of the internal consistency of the samples and their analyses, the individual quality assessment samples must be double-blind and should be labeled as routine samples so that the analyst (and preferably also the laboratory) does not know the relationship between the samples. This reduces the chances of conscious or unconscious efforts to improve the apparent consistency of the analyses.

Table 4. Types of Quality Assessment Samples or Procedures

Double-blind Samples

1. Field Evaluation Samples (FES)

These samples are of known concentration, subjected to the same manipulations as routine samples and introduced in the field at the earliest stage possible. They can be used to detect measurement bias and to estimate precision.

2. Low Level Field Evaluation Samples (LLFES)

These samples are essentially the same as field evaluation samples, but they have very low or non-existent concentrations of the contaminant. They are used for determination of contamination in the sample collection, transport, and analysis processes. They can also be used for determination of the system detection limit (13).

3. External Laboratory Evaluation Samples (ELES)

This sample is similar to the field evaluation sample except it is sent directly to the analytical laboratory without undergoing any field manipulations. It can be used to determine laboratory bias and precision if used in duplicate. We recommend using the same sample as the FES to allow isolation of the potential sources of error. Spiked soil samples have been used as external laboratory evaluation samples in past studies for dioxin, pesticides, and organics (1,6), and natural evaluation samples have been used for metals analysis in soil and liquid samples (3,13).

4. Low Level External Laboratory Evaluation Sample (LLELES)

This sample is similar to the LLFES except it is sent directly to the analytical laboratory without undergoing any field manipulations. It is used to determine method detection limit, and the presence or absence of laboratory contamination. We recommend using the same sample source as for the LLFES to allow isolation and identification of the source of contamination.

5. Field Matrix Spike (FMS)

This is a routine sample spiked with the contaminant of interest in the field. Because of the inherent problems associated with the spiking procedure and recovery it is not recommended for use in field studies (9).

6. Field Duplicate (FD)

An additional sample taken near the routine field sample to determine total within-batch measurement variability. The

differences in the measurements of duplicate and associated samples are in part caused by the short-range spatial variability (heterogeneity) in the soil and are associated with the measurement error in the field crew's selection of the soil volume to be the physical sample (i.e., two crews sent to the same sampling site, or the same crew sent at different times, would be unlikely to choose exactly the same spot to sample).

7. Preparation Split (PS)

After a routine field sample is homogenized, a subsample is taken for use as the routine laboratory sample. If an additional subsample is taken from the routine field sample in the same way as the routine laboratory sample, this additional sample is called a preparation split. The preparation split allows estimation of error variability arising from the subsampling process and from all sources of error following subsampling. This sample might also be sent to a reference laboratory to check for laboratory bias or to estimate inter-laboratory variability. These have also been called replicates (5).

Single-Blind Samples

1. Field Rinsate Blanks (FRB)

These samples, also called field blanks (9), decontamination blanks (14,15), equipment blanks (5), and dynamic blanks (5), are obtained by running distilled, deionized (DDI) water through the sampling equipment after decontamination to test for any residual contamination.

2. Preparation Rinsate Blank (PRB)

These samples, also called sample bank blanks (12,14,15), are obtained by passing DDI water through the sample preparation apparatus after cleaning in order to check for residual contamination.

3. Trip Blank (TB)

These samples are used when volatile organics are sampled, and consist of actual sample containers filled with ASTM Type II water, and are kept with the routine samples throughout the sampling event. They are then packaged for shipment with the routine samples and sent with each shipping container to the laboratory (9). This sample is used to determine the presence or absence of contamination during shipment.

Non-blind Samples

These samples (e.g. Laboratory Control Samples (LCS)) are used in the Contract Laboratory Program to assess bias and precision. For convenience, these samples are described in Appendix E with the definitions being adapted from the CLP Inorganic Statement of Work #788 (10).

Split Samples

Samples can be split to provide:

- samples for both parties in a litigation or potential litigation situation;
- a measure of the within-sample variability;
- materials for spiking in order to test recovery; and
- a measure of the analytical and extraction errors.

The location of the sample splitting determines the components of variance that are measured by the split. A split made in the sample bank (i.e., at the sample preparation facility to which samples are sent from the field) measures error introduced from that level onward. A split made in the field includes errors associated with field handling. A split or series of subsamples made in the laboratory for extraction purposes measures the extraction error and subsequent analytical errors.

Spiked Samples

Spiked samples are prepared by adding a known amount of reference chemical to one of a pair of split samples. Comparing the results of the analysis of a spiked member to that of the non-spiked member of the split measures spike recovery and provides a measure of the analytical bias. Spiked samples are difficult to prepare with soil material itself. Frequently the spike solution is added to the extract of the soil sample. This avoids the problem of mixing, but does not provide a measure of the interaction of the chemicals in the soil with the spike; neither does it provide an evaluation of the extraction efficiency. A predigest spike, as utilized in the CLP (9) would allow a check of the extraction or digestion efficiency. In addition, if the laboratory does the spiking, the spiking is non-blind to the laboratory.

Blanks

Blanks provide a measure of various cross-contamination sources, background levels in the reagents, decontamination efficiency, and other potential error that can be introduced from sources other than the sample. For example, a blank introduced at the earliest point in the field can measure input from contaminated dust or air into the sample. A rinsate blank, i.e., decontamination sample, measures any chemical that may have been on the sampling and sample preparation tools after the decontamination process is completed.

Batch

A batch is defined as a group of samples which are sampled, shipped and analyzed under similar conditions. The total number

of routine and quality assessment samples in a batch is dependent on the desired frequency of quality assessment sampling that a budget will allow. The term batch is synonymous with the term "sample delivery group" as used in the CLP (9).

AVAILABILITY OF QUALITY ASSESSMENT SAMPLES

Presently, performance-evaluation-materials (PEMs) for inorganics in soils are not readily available. (PEMs, closely resemble routine samples, are well characterized, and are provided as unknown samples to a laboratory to demonstrate that the laboratory can produce analytical results within specified limits of performance.) Soil performance-evaluation-materials are available for routine soil characterization for acid deposition purposes (13), and performance materials have been developed for dioxin analysis (1,6). These materials are available as quarterly blind samples; however, adequate PEMs do not exist for analysis of inorganics in hazardous waste soils.

To meet the growing needs for PEMs the Environmental Monitoring Systems Laboratory - Las Vegas (EMSL-LV) in conjunction with EPA's Office of Emergency and Remedial Response (OERR) has begun a project [1] to develop, test, and produce "case-specific" or "site-specific" PEMs. Samples taken at Superfund sites are organized into groups called "Superfund Cases". Each case of samples is sent to a specific CLP laboratory for analysis. The objective is to provide a multi-matrix, multi-analyte, multi-level library or shopping list of PEMs which the Regional site-managers could order by telephone or from a catalog. Each PEM would be included as just another sample within the Case. Ultimately the PEMs would be double-blind to the laboratories. The PEMs would be tailored-made for each Superfund Case of analytical samples to enable more reliable, accurate decision making about Superfund sites.

PEMs are an important component of the rationale that is used to assess variability and bias throughout the measurement process; however, variability and biases may also be assessed without the use of these materials since present availability is limited. An alternative QA design that does not rely on the use of FES and ELES is provided after the discussion of the rationale that is based on the use of PEMs.

---

[1] Butler, L., 1989. Personal communication. Environmental Monitoring Systems Laboratory. Las Vegas, NV.

A RATIONALE FOR ASSESSING ERRORS

Quality Assessment Sample Design

An effective quality assurance program should ensure that the uncertainty associated with the measurement system will be insignificant when compared to the uncertainty allowed for the population of interest. As stated by J.K. Taylor (7):

> "When the uncertainty of a measured value is one-third or less of the permissible tolerance for its use, it can be considered as essentially errorless for that use."

Therefore it is critical that a quality assurance system provides for quantification of total measurement error. Measurement error consists of three major components, i.e. sample collection, preparation, and analysis. Each of these phases can then be divided into smaller components depending on the specific design of the operation.

It is important to realize that if the error associated with the sample collection or preparation phase is large, then the best laboratory quality assurance program is inadequate. Thus a manager should apply the greatest amount of emphasis to the phase that contributes the largest component of error; this will not be possible if the quality assurance design does not provide for error evaluation of the major measurement phases.

The following sample design (Figure 1) is proposed as a complete quality assurance design that will allow determination and control of the various components of measurement bias and precision. It is assumed that only one analytical laboratory is utilized; nevertheless, the design can be applied to multiple laboratories. A multiple-laboratory approach is not discussed here for simplicity. The samples, discussed in the design, were defined previously in Table 4.

Figure 1 depicts how quality assessment samples of several types are treated at the sample collection, preparation, and analysis stages. Starting at the left of the diagram, the collection of a field duplicate is shown. At the location selected for the duplicate, two collocated samples are collected. One is designated as the routine sample (RS), the other as the field duplicate (FD). During the preparation phase, after a routine field sample has been homogenized, a subsample is taken to be used as the routine laboratory sample. If an additional subsample is taken from the routine field sample, this additional sample is called a preparation split. In a similar manner, a subsample is also obtained from a field duplicate to provide the laboratory sample from which a concentration measurement for the

FD will be obtained. These samples are then forwarded to the analytical laboratory for analysis.

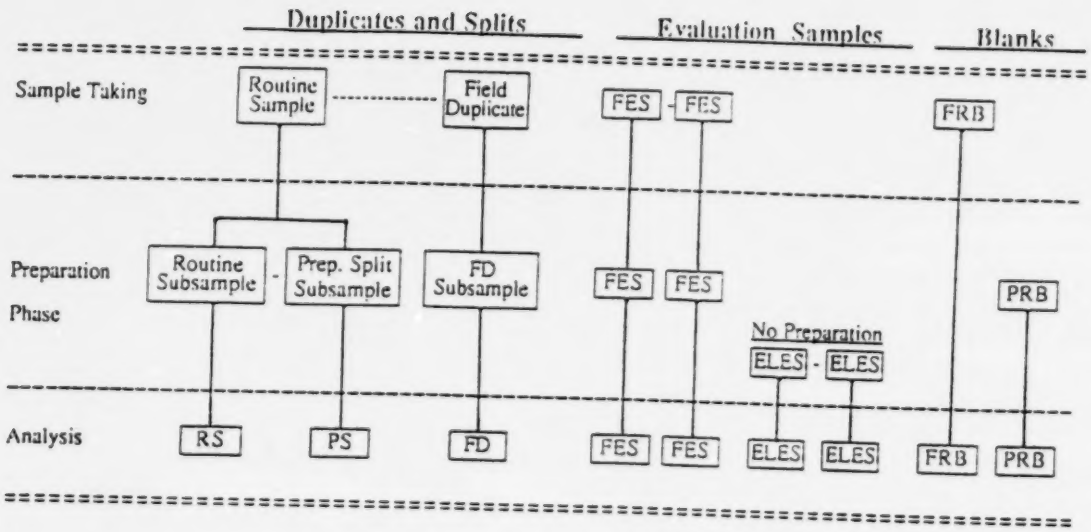
Moving to the middle of the diagram, paired evaluation samples are shown entering the process at two stages. In general, field evaluation samples (FES) are introduced as early in the collection and packaging process as possible. However, in the case of soil sampling, it is normally not possible to pass them over the sampling tools, so they enter the process immediately after the collection step. These samples are then subjected to the same handling and analysis procedures as the other samples. The external laboratory evaluation samples (ELES) are introduced after the preparation stage in such a way that they cannot be identified as QA samples by the laboratory and thereby serve as double-blind samples. These samples are then subjected to the same analytical procedures as routine samples.

On the right side of the diagram, two types of rinsate blanks are shown. The field rinsate blank (FRB) is used to check for sample-collection equipment contamination, and the preparation rinsate blank (PRB) is used to check for preparation equipment contamination.

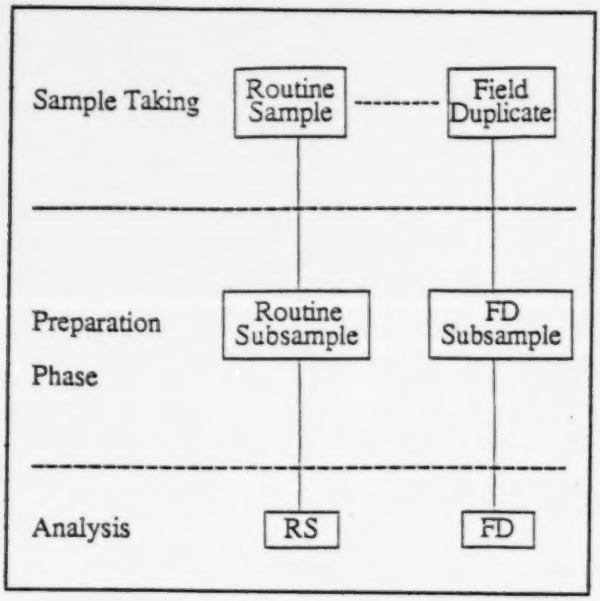
Some consideration must be given to how quality assessment samples should be assigned to batches of routine samples. Each batch should contain either one pair of field evaluation samples or none. Typically, external evaluation samples will only be assigned to batches of samples containing field evaluation samples, and, in such cases, only one pair will be assigned to a batch. Any particular batch may contain zero, one, or several field duplicates and their associated routine samples. However, some attempt should be made to distribute field duplicates throughout the batches from the beginning to the end of the study. Rules for the assignment of preparation splits, and associated routine samples, to batches are the same as for field duplicates. As a general rule, each batch should contain one field rinsate blank and one preparation blank.

Figure 1

QUALITY ASSESSMENT SAMPLES



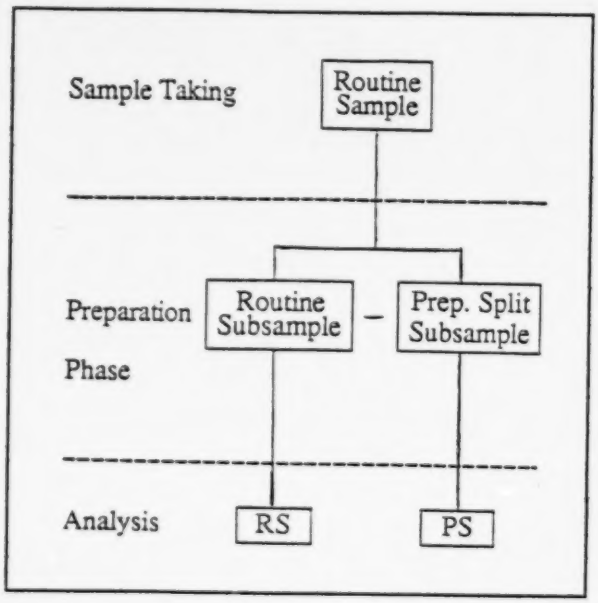
Field Duplicates (FD)



Function of Field Duplicates: To provide data required to estimate total measurement error variance minus between-batch error variance ($\sigma_a^2-\sigma_b^2$). In other words, field duplicates can be used to estimate the sum of all measurement-error variance components except the between-batch error variance component. To assess between-batch errors, field evaluation samples or external lab evaluation samples may be used.

NUMBER REQUIRED:
Since field duplicates are employed in the estimation of total measurement error variance and since an estimate of this variance is required, at least 20 pairs (i.e., routine sample and field duplicate (co-located) sample) or 10 triples (i.e., routine sample and two field-duplicate samples) must be obtained to meet the minimal 20 degrees of freedom objective.

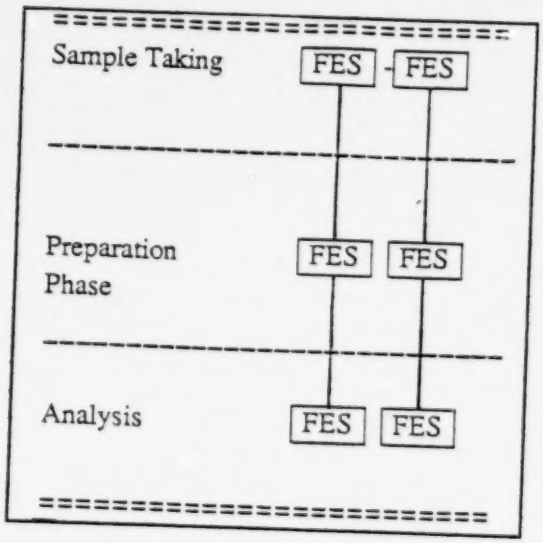
Preparation Splits (PS)



Function of Preparation Splits:
To provide data required to estimate the sum of subsampling and analytical variances ($\sigma^2_{ss}+\sigma^2_a$) To accomplish this, the split must be performed before the sample arrives at the analytical laboratory.

NUMBER REQUIRED:
If the estimation of the components of variance is an important objective of the project, then one should follow the 20 degrees of freedom rule and run at least 20 preparation pairs (i.e., routine subsample and preparation-split subsample). However, unlike estimation of the total measurement error variance, estimation of variance components may be unnecessary in the quality evaluation of some projects. If the estimation of variance components is unnecessary, then the only reason for preparation splits might be for quality control purposes and the number would be determined in terms of the quality control requirements.

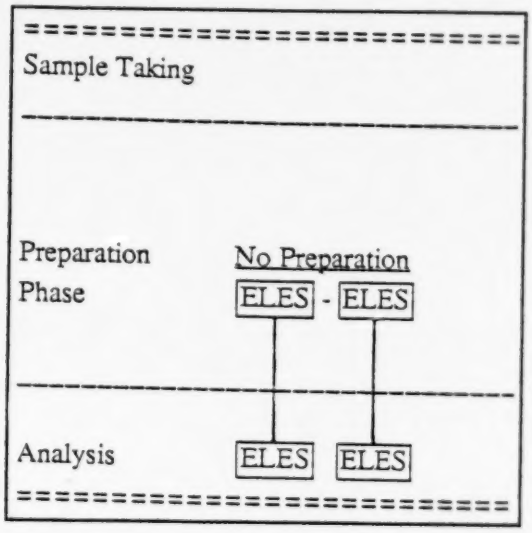
Field Evaluation Samples (FES)



Function of Field Evaluation Samples (in pairs)
To provide data which when taken in conjunction with the data obtained from field duplicate samples and their associated routine samples allows one to obtain unbiased estimates of the total measurement error variance, σ_s^2, the between-batch error variance, σ_b^2, and the sample-collection error variance, σ_{sc}^2. The data from the FES also allow estimation of the handling error variance, σ_h^2, and of the total measurement bias minus the sample collection biases ($B_s - B_s - B_{sc}$).

NUMBER REQUIRED:
Since field evaluation samples are employed in the estimation of the total measurement error variance and since an estimate of this variance is required, at least 21 field-evaluation pairs must be obtained to meet the minimal 20 degrees of freedom for all variance estimates. (The number of required pairs is 21 rather than 20 because one of the variances being calculated from the data is the variance of the paired sample averages; therefore, 21 averages are required to obtain a variance estimate with 20 degrees of freedom.)

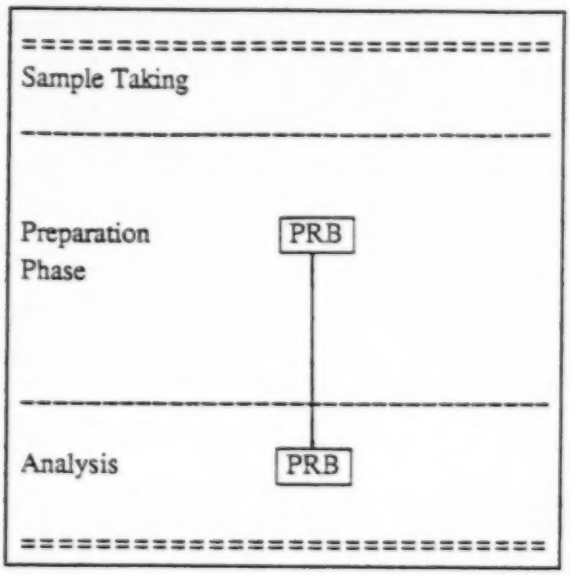
External Laboratory Evaluation Samples (ELES)



Function of ELES:
To provide data required to estimate the sum of the biases due to analysis and to data handling (B_a+B_{ac}), and analytical error variances (σ_a^2).

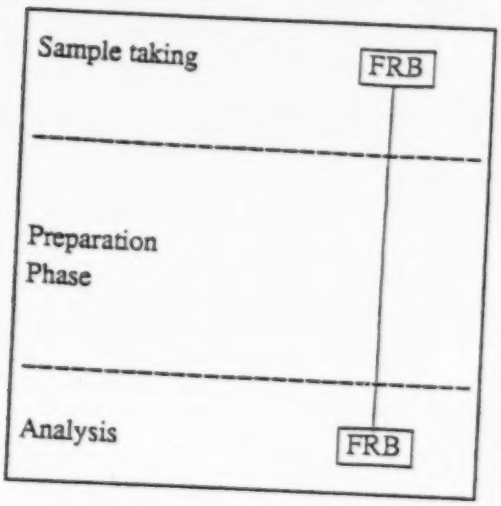
NUMBER REQUIRED:
If the estimation of the components of variance is an important objective of the project, then one should follow the 20 degrees of freedom rule and run at least 20 laboratory evaluation pairs. However, unlike estimation of the total measurement error variance, estimation of variance components and bias components may be unnecessary in the quality evaluation of some projects. If the estimation of variance and bias components is unnecessary, then the only reason for laboratory evaluation samples might be for quality control purposes and the number would be determined in terms of the quality control requirements.

Preparation Rinsate Blanks (PRB)



Function of PRB:
To provide data required to estimate the sum of bias caused by contamination, analysis and data handling ($B_s+B_{sc}+B_{asc}+B_{ac}$).

Field Rinsate Blanks (FRB)



Function of FRB:
To provide data required to estimate the sum of the bias caused by contamination at the time of sampling and at the laboratory and by analysis and data handling ($B_s+B_{sc}+B_{ac}$).

Equations for Estimating Bias and Precision

Once the analytical results are received, computation of bias and precision values is the next step. Bias may be expressed as the difference between the measured concentration of the evaluation samples and the reference or known value for the evaluation sample. A reference value and an expected range of values are usually available for evaluation samples. If the measured values are within this range, then one can say that bias has not been detected. If the measured values are outside this range, then bias may be present and its amount may be estimated from the differences between the measured and the reference values.

Precision is usually described by variance, although standard deviations are sometimes used. However, standard deviations are not additive, while variances are. Table 5 provides equations for the variance estimates. Most of these equations are based on the statistical definitions of variance for the difference between paired values. Subscripts "$_{WFES}$" and "$_{BFES}$" refer to within and between-batch variances, respectively, which are computed from paired field-evaluation samples. Subscripts "$_1$" and "$_2$" refer to individual samples in a pair. In developing these equations, it is assumed that splits and field duplicates were assigned to sample locations such that no location had both a field-duplicate and a preparation split. If duplicates and splits are assigned to the same locations, some of the above variance formulas must be modified. However, all the above variance components can be estimated in either case. The symbol "n" always represents the number of pairs involved. It is also assumed, as will typically be case, that the field evaluation samples are of the same size (weight or volume) as the routine laboratory samples forwarded from the preparation phase: this means that there is no subsampling of the FES in the preparation phase. Triples are not considered here.

Details for computing variance estimates for total measurement error, sample collection, sample handling, subsampling, and analytical error are provided in Table 5. If the estimates of variance components, involving differences of variance estimates, s^2, yield negative values, the reported estimate is zero.

Table 5

EQUATIONS FOR DETERMINING PRECISION AND BIAS

PRECISION

Data Source	Estimate	Parameter Estimated
Field Duplicates	$s^2_{FD} = \sum_{i=1}^{n} [RS_i - FD_i]^2/(2n)$	$(\sigma^2_m - \sigma^2_b) = (\sigma^2_a + \sigma^2_{ss} + \sigma^2_s + \sigma^2_h)$
Prep. Splits	$s^2_{PS} = \sum_{i=1}^{n} [RS_i - PS_i]^2/(2n)$	$(\sigma^2_a + \sigma^2_{ss})$
Field Evaluation Samples	$s^2_{WFES} = \sum_{i=1}^{n} [FES_{1i} - FES_{2i}]^2/(2n)$	$(\sigma^2_a + \sigma^2_h)$
Field Evaluation Samples	$s^2_{BFES} = 2\sum_{i=1}^{n} [\overline{FES_i} - \overline{FES}]^2/(n-1)$ where $\overline{FES_i} = (FES_{1i} + FES_{2i})/2$, and $\overline{FES} = \sum_{i=1}^{n} (FES_{1i} + FES_{2i})/(2n)$	$(\sigma^2_a + \sigma^2_h + 2\sigma^2_b)$
External Lab Eval. Samp.	$s^2_{WLES} = \sum_{i=1}^{n} [ELES_{1i} - ELES_{2i}]^2/(2n)$	σ^2_a
	$s^2_{FD} + (s^2_{BFES} - s^2_{WFES})/2$	σ^2_m
	$s^2_{FD} - s^2_{WFES} - s^2_{PS} + s^2_{WLES}$	σ^2_s
	$(s^2_{BFES} - s^2_{WFES})/2$	σ^2_b
	$s^2_{PS} - s^2_{WLES}$	σ^2_{ss}
	$s^2_{WFES} - s^2_{WLES}$	σ^2_h
	s^2_{WLES}	σ^2_a

Table 5 (continued)

SUGGESTED BIAS FORMULAE

1. Field Evaluation Sample (FES): Bias = 100(F-R)/R %

   where R is the reference value for the FES, and F is the reported measurement of the FES.

2. Field Rinsate Blank (FRB): Bias = 100 X/CRDL %

   where X is the measured value of the FRB, and CRDL is the Contract Required Detection Limit.

3. Preparation Rinsate Blank (PRB): Bias = 100 P/CRDL %

   where P is the measured value of the PRB, and CRDL is as defined above.

4. Pre-Digest Spike: Bias = 100 (SSR-SR-SA)/SA %

   where SSR is the spiked sample result, SR is the sample result, and SA is the spike amount.

5. Post-digest Spike: (same formula as for the pre-digest spike)

AN ALTERNATIVE QA DESIGN THAT DOES NOT EMPLOY FES AND ELES

The quality assurance design given in the preceding section employed field evaluation samples (FES) and external laboratory evaluation samples (ELES). It was assumed that these samples would be double blind samples of homogeneous soil and that the soil would be very similar (i.e., similar in soil type, in concentrations of the pollutants of concern, and in concentrations of other possible chemical interferents) to that to be sampled in the study. This usually implies that a fairly large quantity (or quantities) of soil should be collected from the study site, sent to a laboratory to be dried, mixed, sieved, and split into homogeneous subsamples to be used as FES and ELES. It also requires analysis of a sufficient number of samples by the laboratory to establish the homogeneity of the samples, and the sending of samples to a number of other laboratories to establish a reference value for the FES and ELES. This is a time-consuming process, and the time required for the process may not be available to the RPM prior to the start of the soil sampling study. This section addresses how one may plan the study without FES and ELES so as to still be able to search out bias sources, to estimate some error variance components, and to estimate total measurement error variance.

The basic use of the FES in the preceding section was to estimate between batch variance. As an alternative, it is suggested that additional field duplicates may be employed for this purpose. One may go back to a particular sampling location (e.g., a point at which one sample of soil is taken), and take a fresh (collocated) sample to include with each batch (or with at least 21 randomly selected batches if there are a larger number of batches). If it is difficult to take so many collocated samples from one sampling location, one might use two or three such locations and take collocated samples to include in the batches, alternating between locations (e.g., for two sampling locations A and B, batch 1 has a collocated sample from location A, batch 2 from location B, batch 3 from location A, ...). By comparing the variability between collocated samples that are collected and analyzed in different batches, with the variability within the field-duplicate-and-associated-routine-sample pairs, one can estimate the variability contributed by changes in the measurement process between batches. These collocated samples are actually field duplicates, but because they are used in a different way than the field duplicates encountered in the previous section, they will be identified as batch field duplicates (BFD). Equations for determining the variance estimates using this procedure are given in Table 6. The assumption stated for Table 5 that field duplicates and preparation splits are not associated with the same sampling location is again applied in Table 6. It is shown in Table 6 that the variance component associated handling cannot be separated from that associated with sample collection, and that

Table 6

EQUATIONS FOR DETERMINING PRECISION WITHOUT FES AND ELES

Data Source	Estimate	Parameter Estimated
Field Duplicates	$s^2_{FD} = \sum_{i=1}^{n} [RS_i - FD_i]^2/(2n)$	$(\sigma^2_m - \sigma^2_b) = (\sigma^2_a + \sigma^2_{ss} + \sigma^2_s + \sigma^2_h)$
Prep. Splits	$s^2_{PS} = \sum_{i=1}^{n} [RS_i - PS_i]^2/(2n)$	$(\sigma^2_a + \sigma^2_{ss})$
Batch Field Duplicates[a]	$s^2_{BFD} = \sum_{i=1}^{m} [BFD_i - \overline{BFD}]^2/(m-1)$	σ^2_m
or		
Batch Field Duplicates[b]	$s^2_{BFD} = \sum_{j=1}^{L} \sum_{i=1}^{m_j} [BFD_{ij} - \overline{BFD_j}]^2 / \sum_{j=1}^{L} (m_j - 1)$	σ^2_m
	s^2_{BFD}	σ^2_m
	$s^2_{BFD} - s^2_{FD}$	σ^2_b
	$s^2_{FD} - s^2_{PS}$	$\sigma^2_s + \sigma^2_h$
	s^2_{PS}	$\sigma^2_a + \sigma^2_{ss}$

[a] This equation is appropriate when the m batch field duplicate samples are all taken from one sampling location. $\overline{BFD}$ is the sample mean of the m samples.

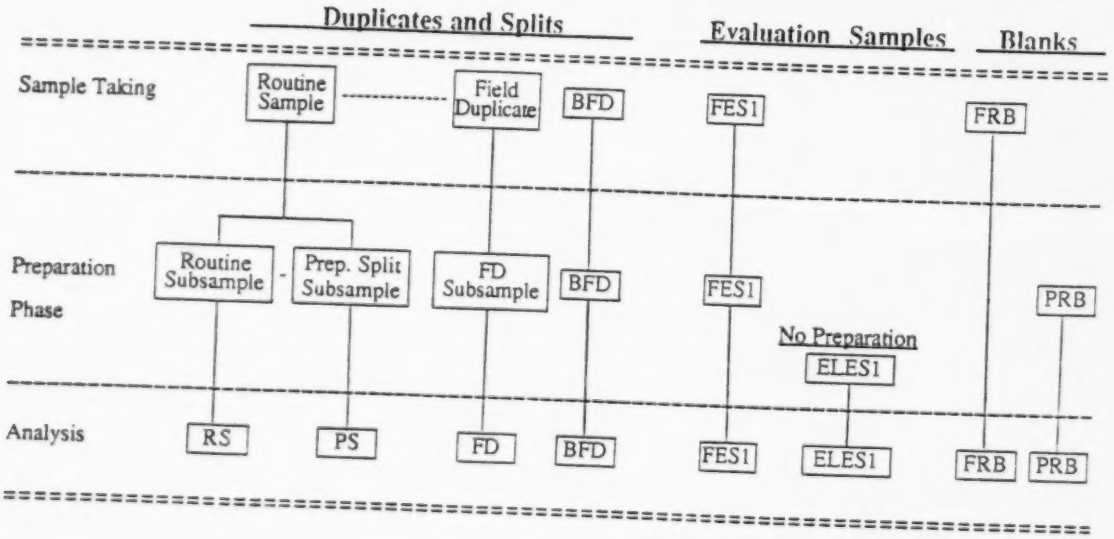
[b] This equation is appropriate when the batch field duplicate samples are from L locations with m_j (>1) BFDs coming from sampling location j. $\overline{BFD_j}$ is the sample mean of the m_j samples taken for location j.

the variance component associated with subsampling cannot be separated from that associated with analysis. This loss of information is a consequence of not using FES and ELES in the study.

The bias detection allowed by use of FES and ELES may be again obtained at least in part by the introduction of well-characterized single-blind samples, containing the contaminants of interest, that are already available from previous studies or from EMSL-LV. These are single-blind samples, since the laboratory analyst will probably be able to distinguish them from the routine samples. These samples will be denoted here by FES1 and ELES1. It will not be necessary to run these samples in pairs as they will not be used in variance estimation. Figure 2 is a diagram of this alternative study that plays the same role as Figure 1 did for the procedure involving FES and ELES.

Figure 2

Alternative Quality Assessment Sample Design



EXAMPLE OF QUALITY ASSESSMENT

The purpose of this example is to show how the guidance in this document can be implemented. Data used in this example were obtained from an actual Superfund site which was contaminated by lead deposition from a smelter; however, arrangement of the data into batches and data from field evaluation samples are fictional and are included for illustrative purposes.

QA/QC from a Pilot Study

A pilot study was conducted over a representative area to determine spatial variability and extent of the lead contamination in order to develop an efficient sampling network for obtaining representative measurements of contamination over a large area. Since measurement variability is known to contribute to the overall variability of data from a sampling effort, a quality assessment program was implemented to assess the variability from the collection, handling, and analysis of the samples. Data from the quality assessment program were intended for use in determining if the measurement variability was so high as to prevent accurate assessments of the spatial variability from being made and whether corrective actions would be required to reduce the measurement variability. One sampling crew and one laboratory were selected for collection and analysis of the samples.

The quality assessment samples are identified in Figure 1 and defined in Table 4. A laboratory control sample was recommended at a rate of one per batch. This sample was obtained from the EMSL-LV EPA laboratory (9), and the acceptable concentration range was provided to the laboratory. This sample was used by the analytical laboratory as a quality control sample; thus, it could not be used to estimate analytical laboratory bias because it was a non-blind sample.

Field evaluation samples were made by sampling 50 kg of a soil type, which was the same as that in the contaminated zone, but was located 5 miles away in an area known from past studies to have background concentrations of lead. The bulk material was then processed as follows:

- the material was air-dried for a one week period
- the material was then sieved, and all material that passed a 2-mm sieve was saved (40 kg of air dried material)
- 16 grams of lead were added to the 40 kg of soil (400 ppm)
- the sample was homogenized by rolling the material in a Teflon-coated drum for 48 hours
- 100 subsamples were made by using a closed-bin riffle splitter
- 10 subsamples, chosen at random, were then shipped to a referee laboratory to check the lead concentrations and to

verify that the lead was equally distributed in each subsample

The total number of samples, by type, utilized in this study were as follows:

routine samples: 180
field evaluation (FES) samples: 6 (3 pairs)
field duplicates (FD): 10
field rinsate blanks (FRB): 10
preparation-split (PS) samples: 10
laboratory control samples (LCS): 10
total samples analyzed: 226
total quality assessment samples: 36
total quality control samples: 10
percentage of QA/QC: 22%

High concentrations of lead were encountered in the field rinsate blank (FRB) from the second batch of samples sent to the analytical laboratory. This problem was not detected until after the 4th batch of samples was sent to the analytical laboratory. Fortunately, this problem was not observed with later batches. Nevertheless, the sampling crew was advised of this problem and told to be more careful. In addition, all samples associated with that batch were resampled and reanalyzed. This problem was not evident with the field evaluation samples (FES) because they could not be used with a split-spoon sampling device. The field evaluation samples were introduced after the sample was taken out of the ground. It was also evident that the contamination did not come from the preparation phase because the preparation blank was acceptable.

Post Pilot Study Data Analysis

After all data were received from the analytical laboratory the equations defined in Table 5 were utilized to calculate estimates of total measurement variance, the sum of sample-collection and sample-handing variances, and between-batch variances.

A computer program, entitled "ASSESS"[2], was developed from the equations in Tables 5 and 6, and data were entered into the program to estimate measurement-error variance components. Data listed in Table 7 were entered into the program to facilitate the calculation of the terms described in Tables 5 and 6. The measured lead concentrations in soil (in mg/kg) are given for 10

[2] This is a public-domain program written in Fortran for use on an IBM PC. It may be obtained by writing to the Exposure Assessment Division, Environmental Monitoring Systems Laboratory, P.O. Box 93478, Las Vegas, NV 89193.

preparation-split pairs and for 10 field-duplicate pairs. The amount of data used has been kept small to make it easier to read and to illustrate the use of the computer program to calculate variances.

The first step is to determine whether a transformation (e.g., taking the natural log (ln) of the values) is needed to stabilize the variance. The estimation of variance components implies that there are unique variances to be estimated that describe measurement-error variance for all measurements. This is not the case if measurement error variances change with sample concentrations. The dependence of measurement error variances on sample concentrations is frequently encountered. Fortunately, this problem can be overcome through the appropriate selection and use of such transformations as are discussed by Hoaglin, et.al. (1983) and Box and Cox (1964). A typical rule of thumb used by statisticians is that if the ratio of the maximum observation to the minimum observation is less than 20, no variance stabilizing transformation is needed; otherwise, the need for a variance stabilizing transformation should be investigated. The information provided by the field-duplicate pairs and the preparation splits is useful in deciding whether a transformation is required to stabilize variance with respect to sample concentration.

For each field-duplicate pair and each preparation-split pair in Table 7, the sum and absolute difference of the two measurements is calculated. One compares how the pair absolute differences change as the pair sums change, which is equivalent to comparing how the sample standard deviations change as the associated sample means change. For the field-duplicate pairs, one observes that the differences associated with the larger sums tend to be larger than those associated with smaller sums. For example, the median difference associated with the five largest sums is 96, while the median difference associated with the five smallest sums is only 28. Similarly, for the preparation-split pairs, one finds the median difference associated with the five largest sums is 28, while the median difference associated with the five smallest sums is only 1.3. This is reasonably clear evidence measurement error variances are changing with sample concentrations and that a transformation is required. There is insufficient information available in the table to choose an appropriate variance stabilizing transformation. However, lead concentration data from other soil sampling studies indicate that the simple logarithmic transformation (Y=ln(lead concentration)) satisfactorily stabilizes the variances. The log-transformation was performed on the data in Table 7, and the results are given in Table 8 along with the variance component calculations and estimates from the ASSESS program.

Table 7. Quality Assessment Data

QUALITY EVALUATION DATA[3]				Transformed? No						
Batch	RS	FD	PS	FES Pairs		ELES Pairs	(RS+FD)/2[4]	|RS-FD|[5]	(RS+PS)/2	|RS-PS|
1				448.0	505.0		.0	.0	.0	.0
4				475.0	488.0		.0	.0	.0	.0
8				423.0	424.0		.0	.0	.0	.0
1	389.0		430.0				.0	.0	409.5	41.0
1	246.0	410.0					328.0	164.0	.0	.0
2	33.4		32.1				.0	.0	32.8	1.3
2	960.0	780.0					870.0	180.0	.0	.0
3	221.0		244.0				.0	.0	232.5	23.0
3	180.0	208.0					194.0	28.0	.0	.0
4	60.0		72.0				.0	.0	66.0	12.0
4	87.0	221.0					154.0	134.0	.0	.0
5	275.0		233.0				.0	.0	254.0	42.0
5	349.0	400.0					374.5	51.0	.0	.0
6	474.0		446.0				.0	.0	460.0	28.0
6	478.0	382.0					430.0	96.0	.0	.0
7	33.5		32.7				.0	.0	33.1	.8
7	33.0	33.3					33.2	.3	.0	.0
8	1,360.0		1,340.0				.0	.0	1,350.0	20.0
8	104.0	128.0					116.0	24.0	.0	.0
9	313.0		294.0				.0	.0	303.5	19.0
9	201.0	161.0					181.0	40.0	.0	.0
10	67.0		67.0				.0	.0	67.0	.0
10	275.0	199.0					237.0	76.0	.0	.0

[3] Concentrations in mg/kg

[4] The average concentration of the routine sample and field duplicate is computed for the purpose of determining whether a transformation of the data is required.

[5] This is computed to determine whether a transformation of the data is required and is equal to the standard deviation times the square root of 2.

Table 8. Transformed Quality Assessment Data

QUALITY EVALUATION DATA[6]				Transformed? ln						
Batch	RS	FD	PS	FES Pairs		ELES Pairs	(RS+FD)/2[7]	|RS-FD|[8]	(RS+PS)/2	|RS-PS|
1				6.105	6.225		.000	.000	.000	.000
4				6.163	6.190		.000	.000	.000	.000
8				6.047	6.050		.000	.000	.000	.000
1	5.964		6.064				.000	.000	6.014	.100
1	5.505	6.016					5.761	.511	.000	.000
2	3.509		3.469				.000	.000	3.489	.040
2	6.867	6.659					6.763	.208	.000	.000
3	5.398		5.497				.000	.000	5.448	.099
3	5.193	5.338					5.265	.145	.000	.000
4	4.094		4.277				.000	.000	4.186	.182
4	4.466	5.398					4.932	.932	.000	.000
5	5.617		5.451				.000	.000	5.534	.166
5	5.855	5.991					5.923	.136	.000	.000
6	6.161		6.100				.000	.000	6.131	.061
6	6.170	5.945					6.058	.224	.000	.000
7	3.512		3.487				.000	.000	3.499	.024
7	3.497	3.506					3.501	.009	.000	.000
8	7.215		7.200				.000	.000	7.208	.015
8	4.644	4.852					4.748	.208	.000	.000
9	5.746		5.684				.000	.000	5.715	.063
9	5.303	5.081					5.192	.222	.000	.000
10	4.205		4.205				.000	.000	4.205	.000
10	5.617	5.293					5.455	.323	.000	.000

Total measurement error variance	.077
Sample collection variance	Insufficient samples for the computation to be made
Between batch variance	.004
Subsampling variance	Insufficient samples for the computation to be made
Handling variance	Insufficient samples for the computation to be made

[6] Concentrations in mg/kg

[7] The average concentration of the routine sample and field duplicate is computed for the purpose of determining whether a transformation of the data is required.

[8] This is computed to determine whether a transformation of the data is required and is equal to the standard deviation times the square root of 2.

Figure 3 from the ASSESS program further illustrates the need for a transform of the original data. The standard deviation of the data from routine samples and field duplicates increases with the average concentration. A logarithmic transform of the data will stabilize the standard deviation of the data over the measured concentration range. After this has occurred, the variances may be computed for the purpose of assessing variability throughout the measurement process. The ASSESS program allows these calculations to be performed easily.

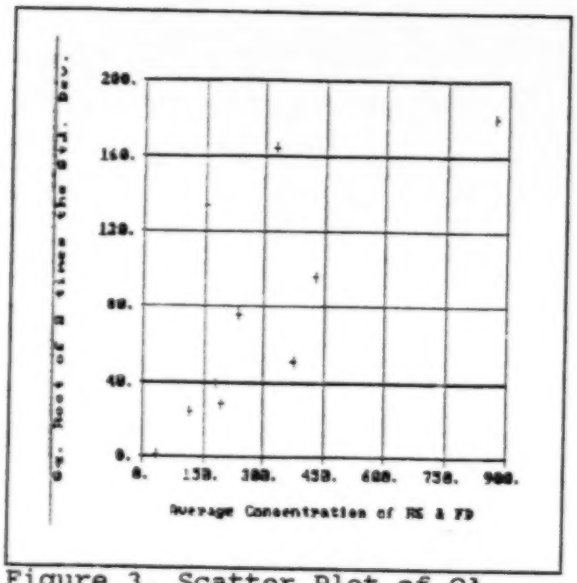


Figure 3. Scatter Plot of QA Data from ASSESS

The variances reported in Table 8 indicate that the sum of the variances arising from sample collection and from sample handling amount to about 7/8ths of the total measurement-error variance. The estimate of the sum of the variances caused by sample collection and sample handling is given by (see Table 5)

$$s_{FD}^2 - s_{pa}^2 = 0.0730 - 0.0045 = 0.0685,$$

while the estimate of total measurement-error variance is given by

$$s_{FD}^2 + (s_{BPES}^2 - s_{WPES}^2)/2 = 0.0730 + (0.0100 - 0.0025)/2 = 0.0768$$

However, the estimate s_{BPES}^2, based on only 2 degrees of freedom (3 pairs -1), may underestimate the true variance σ_{BPES}^2 by a factor of 30 or more (Table 3). The other estimates, s^2, of variance are based on 10 or fewer degrees of freedom and may underestimate the true variances by factors of 3 or more. If all of the estimates of variance had been based on at least 20 degrees of freedom each, one would have much more confidence in the estimates and would certainly be justified in instituting a more rigorous training program for the sample-taking crews and in considering an increase in the volume of soil in each sample. In point of fact, the sum of the variances caused by sample collection and sample handling was such a large portion (approximately 1/3) of the total (measurement plus spatial) variation that action was taken to reduce its contribution in the primary study. While more data from the use of field evaluation samples and external laboratory evaluation samples would have

been useful in implementing the rationale for assessing errors and variability in all phases of the pilot study, sufficient information was provided from the existing quality assessment samples to begin making some changes. Tighter adherence to the rationale during the main part of the study would ensure that sufficient data were available to accurately assess the significance and sources of variability during the study of the entire site.

Figure 4 from the ASSESS program illustrates the range in which the estimates of the various variance components can be expected to occur within a 95% confidence interval. It is clear from the length of the line for S_{BTES}^2 that greater use of field evaluation samples would have improved the assessment of between-batch variability as well as of the total measurement error and sample collection variances.

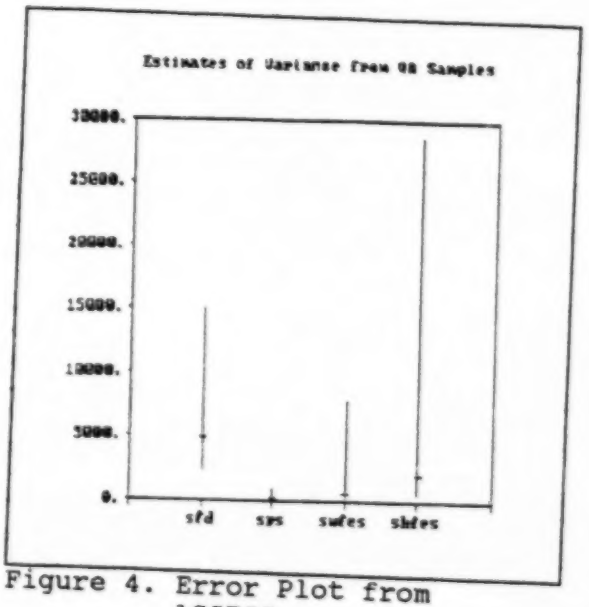


Figure 4. Error Plot from ASSESS

Development of the QA Plan for the Primary Study

The DQOs for the primary study were developed with the background data from the pilot study. Goals were established for accuracy, precision, completeness, comparability and representativeness of the data to be collected during the expanded study. Historical data on variability in major portions of the measurement process were input into the ASSESS program for reference. Based upon the limited number of FES, lack of ELES in the pilot study and the importance of having reliable assessments of data quality throughout the measurement process, Table 3 was used to determine the added number of QA samples that were required to better estimate variability during sample collection, handling, transportation, subsampling and analysis. Even though the personnel, procedures, and analytical equipment would be identical in the primary study to that in the pilot study, the decision-makers felt that 20 degrees of freedom were needed for all quality assessment samples to permit assessments of variability to be within a factor of roughly two to the actual value, at least to the 95% confidence interval. These assessments would confirm that the changes made during the pilot study were effective in reducing sampling and measurement variability to acceptable levels, i.e., to permit spatial variability to be accurately assessed.

RECOMMENDATIONS FOR FUTURE WORK

1. Greater attempts to define and standardize QA/QC terms need to be made in conjunction with the Quality Assurance Management Staff (QAMS) within the Office of Research and Development (ORD) and the program offices.

2. Protocols and materials for the preparation of QA/QC samples for the field need to be reviewed further and described in greater detail. In some cases, new materials and protocols will need to be developed and standards established.

3. The rationale presented in this document needs to be developed further to integrate it with the work of QAMS. Data quality objectives (DQOs) are important in determining the level of QA/QC for a study, and QAMS effort to develop a standardized approach to the development of DQOs through the use of computer software could incorporate the rationale presented in this document. It appears that this rationale could be translated into a spreadsheet or expert systems computer program.

4. Greater characterization of commonly used sampling methods needs to be made. The choice of a sampling method determines to some extent the amount of QA/QC involved in a study. For methods such as the portable x-ray fluorescence instrument, the volume of earth sampled, minimum detection level, interferences, and range of contaminants detected are some of the characteristics that need to be defined, when practicable, on a scale common to the other sampling methods.

5. The rationale presented in this document needs to be evaluated at several actual Superfund site investigations. If the rationale proves to be workable and worthwhile, the rationale needs to be adopted for use at all Superfund site investigations to try to achieve a uniform measure of data quality from all of the investigations of inorganics in soil.

6. Training may be conducted prior to a study or during a study. The optimum approach is to complete the training and evaluation of sampling crews prior to the initiation of a major study. The feasibility of establishing a national training/certification program for sampling crews should be considered further.

7. Specific sub-sampling techniques need to be defined and developed for utilization in the field and the laboratory.

REFERENCES

(1) Bennet, Tom. 1989. Standard Operating Procedures. Region IV Analytical Services Branch, U.S. Environmental Protection Agency.

(2) Box, G.E.P., and D.R. Cox. 1964. The Analysis of Transformations. Journal of the Royal Statistical Society, Series B. Vol. 26(2):211-243.

(3) Drouse, S. K., D.C. Hillman, L.W. Creelman, and S. J. Simon, 1986. National Surface Water Survey, Eastern Lake Survey (Phase 1 - Synoptic Chemistry) Quality Assurance Plan. EPA/600/4-86/008. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

(4) Hoaglin, D.C., F. Mosteller, and J.W. Tukey. 1983. Understanding Robust and Exploratory Data Analysis. John Wiley & Sons. New York, NY

(5) La Barge, R. 1989. "A Programmatic Approach to Achieving Data Quality Objectives." presented at 2nd. Annual Symposium for Quality Assurance for Ecological Monitoring, Feb. 1989. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

(6) Morey, Debra. 1989. "Standard Operating Procedures. U.S. Environmental Protection Agency, Region VIII.

(7) Taylor, J.K. 1987. "Quality Assurance of Chemical Measurements"

(8) U.S. DOE. 1987. The Environmental Survey Manual. DOE/EH-0053.

(9) U.S. EPA. 1987. Data Quality Objectives for Remedial Response Activities - Development Process. EPA/540/G-87/003.

(10) U.S. EPA. 1988. USEPA Contract Laboratory Program - Statement of Work for Inorganics Analysis - Multi-Media Multi Concentration. SOW No. 788.

(11) U.S. EPA. 1988. Sampling for Hazardous Materials. Environmental Response Team, Office of Emergency and Remedial Response, Hazardous Response Support Division, Washington, D.C.

(12) U.S. EPA. 1980. Quality Assurance Plan Love Canal Study. LC1-619-026, EPA Contract 68-02-3168.

(13) U.S. EPA. 1989. Direct/Delayed Response Project: Quality Assurance Plan for the Mid-Appalachian Region of the United States (Draft). Environmental Monitoring Systems Laboratory, Las Vegas, Nevada.

(14) U.S. EPA. 1984. Soil Sampling Quality Assurance Users Guide (1st. Edition). Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. EPA/600/4-84-043

(15) U.S. EPA. 1989. Soil Sampling Quality Assurance Users Guide (2nd. Edition). Environmental Monitoring Systems Laboratory, Las Vegas, Nevada. EPA 600/8/89/046

APPENDIX A

KEY COMPONENTS

A rigorous program to assess the quality of data cannot be developed and implemented if key components for a field study are neglected. Those components include training, pilot studies, a variety of audits to assess the effectiveness of the QA/QC program, and documentation.

Training

Training is an integral part of an effective quality assurance program. It should furnish the essential knowledge needed by all study participants to assure that plans, methods, and procedures are all accomplished as designed. The training program should review key principles and point out changes in protocols to experienced workers. At the same time it should orient the new employee to all the study methods. Training should always occur at the earliest possible time before a study in order to give personnel time to make adjustments. However, training typically occurs, in varying degrees, throughout multi-phased studies as personnel change and more is learned about the site, methods and procedures. Training may be formal, or informal, in the classroom, or on the job. Training should include lectures on sampling principles, a demonstration of procedures, question and answer sessions, and hands-on sampling practice.

A practical way to implement a training program is to integrate it into a "preliminary" study. However, additional training may also occur during a "pilot" study as personnel are evaluated on the methods and procedures that are expected to be used during the main study.

Pilot Study

Pilot studies may serve as the impetus for further training before the full-scale study. The purpose of a pilot study is to evaluate the logistics, equipment, sampling plans and analytical protocols prior to implementation of the full study. It should also provide a test of the study design, quality assurance design, and data interpretation plan. Pilot studies are recommended for all programs so that state-of-the-art measurement methods and study designs can be fully tested before the full study begins. In order to be useful, the pilot study should employ all of the plans for the full-scale project, including the same personnel, management structure, equipment, and procedures. After the pilot study is complete, the study methods can be carefully assessed to see whether or not changes need to be made before the full-scale study starts. The data must be interpreted so that the study methods may be evaluated in relation to how the

study objectives are being met. It is critical that managers provide time and resources to evaluate the pilot study and incorporate changes before the start of the next phase of a site investigation.

The guidance in this document, to assess errors in field sampling, may be used to assess the proficiency of a training effort by allowing the measured errors to be compared against stated data quality objectives.

Audits

An adequate QA program ensures that the quality of the final product meets the DQOs. Audits are an integral part of the QA process and are vital for assuring that program procedures are being implemented.

Audits are performed to document the implementation of the quality assurance program plan, quality assurance project plan and/or associated operational protocols. Four specific kinds of audits can be used to determine the status of the measurement systems, the adequacy of the data collection systems, the completeness of documentation of data collection activities and the abilities of the program management to meet the mandated data collection and data quality objectives. These four audit types are respectively, Performance Audits, Technical System Audits, Data Quality Audits and Management System Audits.

* Performance Audits (PA) are generally based on Quality Assessment or Evaluation (QE) samples. Samples having known concentrations may be tested as unknowns in the laboratory or a sample may be analyzed for the presence of certain compounds. Performance audits are used to determine objectively whether an analytical measurement system is operating within established control limits at the time of the audit.

* Technical System Audits (TSA) are qualitative on-site audits that evaluate the technical aspects of field operations against the requirements of the approved protocols and QA plans. TSA reports will note any problems, allowing corrective action to be taken to protect the validity of future data.

* Data Quality Audits (DQA) are evaluations of the documentation associated with data quality indicators of measurement data to verify that the generated data are of known and documented quality. This is an important part of the validation of data packages showing that the methods and SOPs designated in the QA plans were followed, and that the resulting data set is a functional

part of satisfying the established DQOs. The results are vital to decisions regarding the legal defensibility of the data should it be challenged in litigation.

* A Management System Audit (MSA) is a formal review of an entire program, e.g., a review of a state's QA program, or a review of a state-contracted Laboratory. In a MSA, key elements in the program, e.g., lab certification program, QC in field operations, and QC in the certified lab, are evaluated to see if QA is being implemented. If deficiencies are detected, corrective actions are suggested and implementation monitored.

The guidance in this document is particularly useful for the performance audit.

Documentation

Documented procedures should be developed prior to a study and followed. Errors can increase and blunders can occur in any measurement program through inadequately prepared and reviewed documentation. Data transcribed onto paper or recorded on magnetic media must be checked for accuracy on a timely basis by qualified personnel. Measures to assess and minimize errors, as described in this document, are not going to be effective if adequate documentation is not developed and reviewed on a timely basis.

It is important that field personnel follow specified, documented procedures. If changes in program execution and design (e.g., sample site selection, number of samples to be collected, sampling intervals, and tools) are required, these changes, with appropriate rationale, must also be documented.

APPENDIX B

DEFINITIONS

Quality Assurance

A system of activities whose purpose is to provide to the producer or user of a product or service the assurance that it meets defined standards of quality. It consists of two separate, but relate activities, quality control and quality assessment.

Quality Control

The overall system of activities whose purpose is to control the quality of the measurement data so that they meet the needs of the user.

Quality Assessment

The overall system of activities that provide an objective measure of the quality of data produced.

Soil

The soil referred to in this document encompasses the mass (surface and subsurface) of unconsolidated mantle of weathered rock and loose material lying above solid rock. Further, a distinction must be made as to what fraction of the unconsolidated material is soil and what fraction is not. The soil component here is defined as all mineral and naturally occurring organic material that is 2 mm or less in size. This is the size normally used to distinguish between soils (consisting of sands, silts, and clays) and gravels. In addition, the 2-mm size is generally compatible with analytical laboratory methods, capabilities, and requirements.

The non-soil fraction (e.g., automobile fluff, wood chips, various adsorbents and mineral/organic material greater than 2-mm in size) must also be addressed in sampling and monitoring. This fraction may contribute and/or contain a greater amount of contaminant(s) than the associated soil fraction. At sites in which this occurs, reporting contaminant levels only in the soil fraction will ultimately lead to inappropriate and incorrect decision making. Decision makers must realize that a number of problems are normally encountered in obtaining and using data from the non-soil components. For example, questions arise concerning the validity of data obtained from the analysis of materials that do not meet the size and volume requirements for which the analytical processes were validated. Also, standard reference and audit materials are not available to substantiate and validate the analytical results.

The current recommended procedures are to identify and record the type and volume of non-soil material for each sample collected with a minimum of 10 percent (%) of these non-soil samples submitted for analysis. Data from the non-soil material are important to the assessment of the representativeness of the soil sampling/monitoring program. The behavior of contaminants in the soil environment is a function of the contaminant's and soil's physical and chemical properties. Soil sorption (the retention of substances by adsorption or absorption) is related to properties of the contaminant (e.g., solubilities, heats of solution, viscosity, and vapor pressure) and to properties of soils (e.g., clay content, organic content, texture, permeability, pH, particle size, specific surface area, ion exchange capacity, water content, and temperature). The soil components that are most associated with sorption are clay content and organic matter. The soil particle surface characteristics thought to be most important in adsorption are surface area and cation exchange capacity (CEC).

Standard Additions

A procedure called <u>standard additions</u> is commonly used to detect bias in chemical analysis. In this procedure, known amounts of standard solutions are added to aliquots of soil samples. It is recommended that this be done in the field or in a field laboratory. The main problem encountered is that mixing soils to obtain homogeneity is difficult in a laboratory, and even more so in the field. Several known quantities of the standard are added to the aliquots of the soil samples. The analytical results should follow a straight line:

$$y = a + bx,$$

where x is the increase in concentration caused by the addition and y is the value obtained by the laboratory. Bias is indicated if the data do not follow a straight line, or if $a < 0$. If the units of x and y are the same, the value of b should be near one, and a significant deviation from one would indicate a proportional bias.

APPENDIX C

Soil Sampling Methods Table

Type of Sampler	Obtains Core Sample		Most Suitable Core Types			Operation in Stoney Soils		Most Suitable Soil Moisture Conditions			Access to Sampling Sites During Poor Soil Conditions		Relative Sample Size		Sampling Depth			Labor Requirements		Cost	
	Yes	No	Cho[a]	Cho'less[b]	Either	Fav	Unfav	Wet	Dry	Either	Yes	No	Small	Large	Shallow	Deep	Either	1	2/More	Small	Large
I. Hand-Held																					
Spoons		x			x	x				x	x			x	x				x		x
Scoops		x			x	x				x	x			x	x				x		x
Shovels		x			x	x				x	x			x			x		x		x
Augers																					
Screw-type Augers		x			x		x	x			x		x				x	x			x
Barrel Augers																					
Post-hole Auger		x	x			x		x			x			x			x				
Dutch Auger		x	x			x		x						x			x				
Regular Barrel Auger		x	x			x											x				
Sand Augers		x		x		x			x		x		x				x				
Mud Augers		x	x			x		x			x			x			x				
Tube-Type Samplers																					
Soil Probes																					
Wet Tips	x				x		x	x			x						x				
Dry Tips	x				x		x	x			x		x				x	x	x		
Veihmeyer Tubes	x				x					x			x			x					
Finn-nailed Tube Samplers	x		x				x			x	x			x	x						
Peat Samplers	x		x				x	x			x			x			x				
II. Power Driven Auger																					
Hand-Held Screw Type		x			x	x				x	x			x			x		x		x
Power Auger					x							x				x					
Truck Mounted Auger																					
Tripod Mounted Drive Sampler	x		x				x			x		x		x			x				
Split Spoons	x		x				x			x		x		x			x				

a. Cohesive Soils (e.g., clay)
b. Cohesiveless soils (e.g., dry sand)

APPENDIX D

THE LOGNORMAL DISTRIBUTION AND LOGARITHMIC TRANSFORMATIONS

If the random variable $W = Ln(X)$ has a normal distribution with mean μ_w and variance σ_w^2 (i.e., $N(\mu_w,\sigma_w^2)$), then the random variable X has a lognormal distribution (Johnson and Kotz, 1970, Chapter 14[9]) with mean

$$E(X) = \mu_x = \exp(-\mu_w\sigma_w + \sigma_w^2/2),$$

with variance

$$V(X) = \sigma_x^2 = \mu_x^2(\exp(\sigma_w^2) - 1)$$

and with standard deviation

$$\sigma_x = \mu_x \sqrt{(\exp(\sigma_w^2)-1)}.$$

Note that the relative standard deviation (rsd) for X is $\sqrt{(\exp(\sigma_w^2)-1)}$ which is independent of the mean, μ_x, while the standard deviation is a linear function of μ_x. (The logarithmic transformation is used to stabilize measurement variance relative to concentration when the data indicates that the standard deviation is a linear function of the sample concentration.) Further, note that the series expansion about zero of the square of the rsd as a function of σ_w^2 is:

$$[rsd(X)]^2 = \sigma_w^2 + (1/2!)(\sigma_w^2)^2 + (1/3!)(\sigma_w^2)^3 + \ldots$$

and so when σ_w^2 is a number near zero, $[rsd(X)]^2 \approx \sigma_w^2$. The square of the rsd is sometimes called the rel-variance.

---

[9] Johnson, N.L., and S. Kotz. Continuous Univariate Distributions -1. Houghton Mifflin Co., Boston, MA 1970 300 pp.

APPENDIX E

NON-BLIND
QUALITY ASSESSMENT SAMPLES IN THE CONTRACT LABORATORY PROGRAM

1. Laboratory Control Sample (LCS) - A sample of well-characterized soil, whose analyte concentrations are known to the laboratory, is used for internal laboratory control (10,15). This sample is also called a quality control audit sample (13).

2. Pre-digest Spike Sample - A routine sample in which a known quantity of analyte is added to an aliquot of the sample. It is used to determine bias from the digestion and analysis of components.

3. Post-digest Spike Sample - A routine sample in which a known quantity of analyte is added to an aliquot after the digestion process is completed. It is used to determine bias from the analytical or detection phase. When used in combination with a pre-digest spike sample, the bias from the digestion phase may be determined by difference.

4. Analytical Laboratory Duplicate (ALD) - This sample is a subsample of a routine sample which is analyzed by the same method. It is used to determine method precision, but because it is a non-blind sample, or known to the analyst, it can only be used by the analyst as an internal control tool and not as an unbiased estimate of analytical precision.

5. Initial Calibration Verification (ICV) and Continuing Calibration Verification (CCV) Solutions - These are prepared solutions containing known concentrations of analytes that originated from a different source as the calibration standards. They are used as an independent check of the instrument calibration accuracy. The CCV samples are normally run in an ordered fashion after a specified number of routine samples.

6. Initial Calibration Blank (ICB) and Continuing Calibration Blank (CCB) Solution. These are blank samples run at the same frequency as the ICV and CCV, and they are used to check for instrument baseline drift.

7. CRDL Standard for ICP and AA - This is a solution standard at a concentration of two times the CRDL, or two times the IDL, whichever is greater. It is used during each run in place of a formal instrumental detection limit determination to assure the instrument is running properly.

8. Linear Range Verification Check Standard - This standard is

a solution of known analyte at concentrations within the upper limit of the linear range. Above this range, the samples must be diluted.

9. ICP Interference Check Sample - This sample contains two parts. Part A contains potential interfering analytes, and Part B contains both the analytes of interest and the target analytes. Part A and B are analyzed separately to determine the potential for interferences.

APPENDIX F

Upper Confidence Limits for the Variance, σ^2, as a Function of the Number of Degrees of Freedom for the Variance Estimate, s^2.

Degrees of Freedom	Levels of Confidence (%) 90	95	99
2	$9.49s^2$	$19.49s^2$	$99.50s^2$
3	$5.13s^2$	$8.52s^2$	$26.13s^2$
4	$3.76s^2$	$5.63s^2$	$13.46s^2$
5	$3.10s^2$	$6.01s^2$	$9.02s^2$
6	$2.72s^2$	$3.67s^2$	$6.88s^2$
7	$2.47s^2$	$3.23s^2$	$5.65s^2$
8	$2.29s^2$	$2.92s^2$	$4.86s^2$
9	$2.16s^2$	$2.71s^2$	$4.31s^2$
10	$2.05s^2$	$2.54s^2$	$3.91s^2$
11	$1.97s^2$	$2.40s^2$	$3.60s^2$
12	$1.90s^2$	$2.29s^2$	$3.36s^2$
13	$1.85s^2$	$2.21s^2$	$3.17s^2$
14	$1.80s^2$	$2.13s^2$	$3.00s^2$
15	$1.76s^2$	$2.07s^2$	$2.87s^2$
16	$1.72s^2$	$2.01s^2$	$2.75s^2$
17	$1.69s^2$	$1.96s^2$	$2.65s^2$
18	$1.66s^2$	$1.92s^2$	$2.57s^2$
19	$1.63s^2$	$1.87s^2$	$2.49s^2$
20	$1.61s^2$	$1.84s^2$	$2.42s^2$
21	$1.59s^2$	$1.81s^2$	$2.36s^2$
22	$1.57s^2$	$1.78s^2$	$2.31s^2$
23	$1.55s^2$	$1.76s^2$	$2.26s^2$
24	$1.53s^2$	$1.73s^2$	$2.21s^2$
25	$1.52s^2$	$1.71s^2$	$2.17s^2$
30	$1.46s^2$	$1.62s^2$	$2.01s^2$
40	$1.38s^2$	$1.51s^2$	$1.80s^2$
50	$1.33s^2$	$1.44s^2$	$1.68s^2$
100	$1.21s^2$	$1.28s^2$	$1.43s^2$